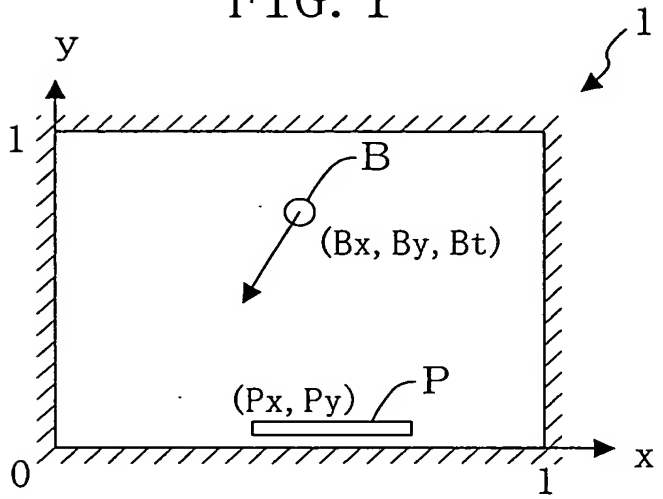
FIG. 1



Mp(LEFT, RIGHT, STAY)

FIG. 2

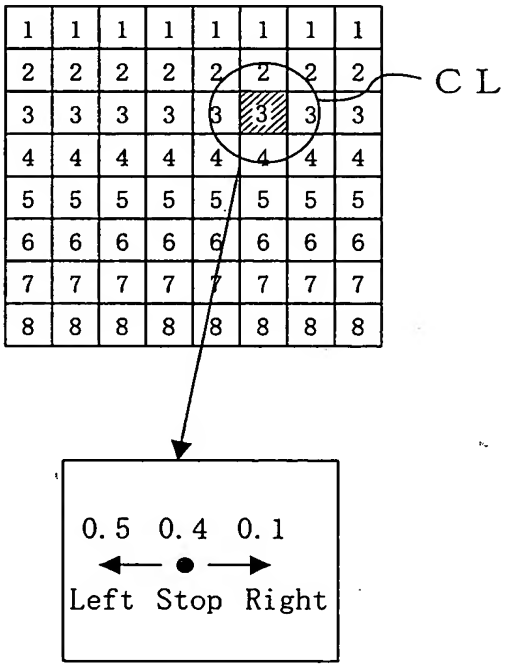


FIG. 3

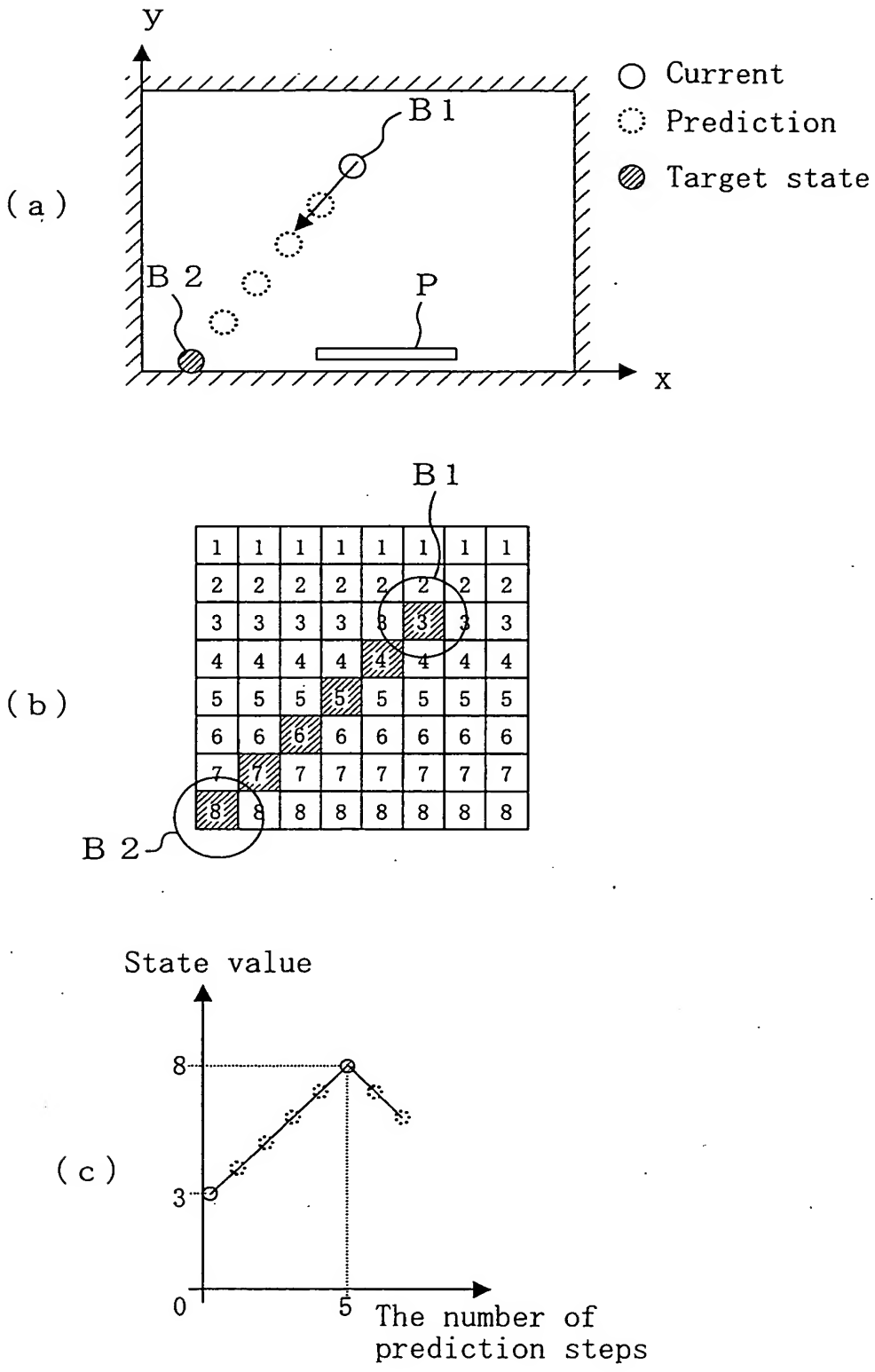

(a)
y
x
B 1
B 2
P
Current
Prediction
Target state
(b)
B 1
B 2
(c)
State value
8
3
0
5
The number of
prediction steps

FIG. 4

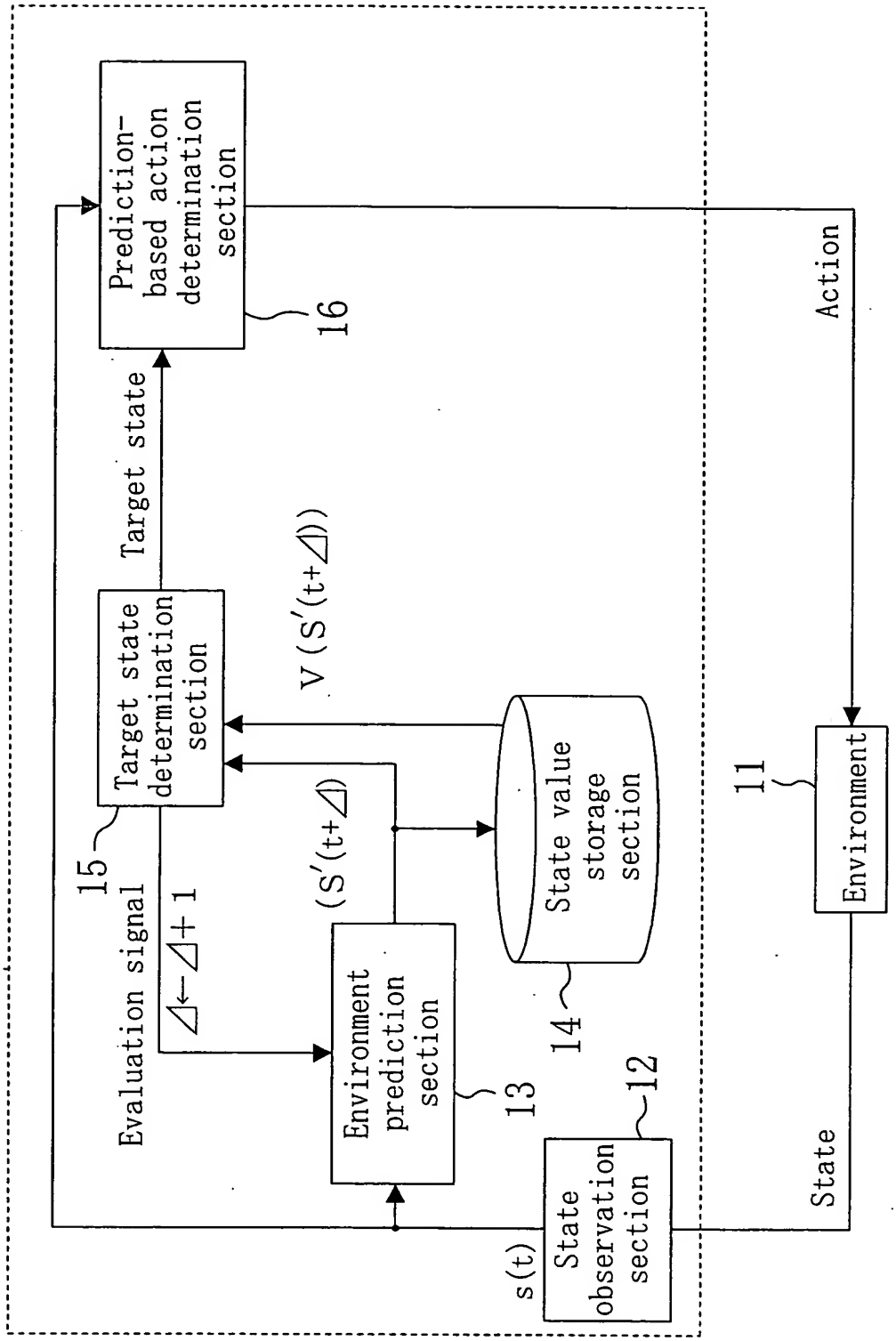

10
Prediction-based action determination section
16
Action
Target state
Target state determination section
15
V (S'(t+Δ))
Evaluation signal
Δ←Δ+1
(S'(t+Δ)
State value storage section
14
Environment prediction section
13
12
11
Environment
State observation section
s(t)
State

FIG. 5

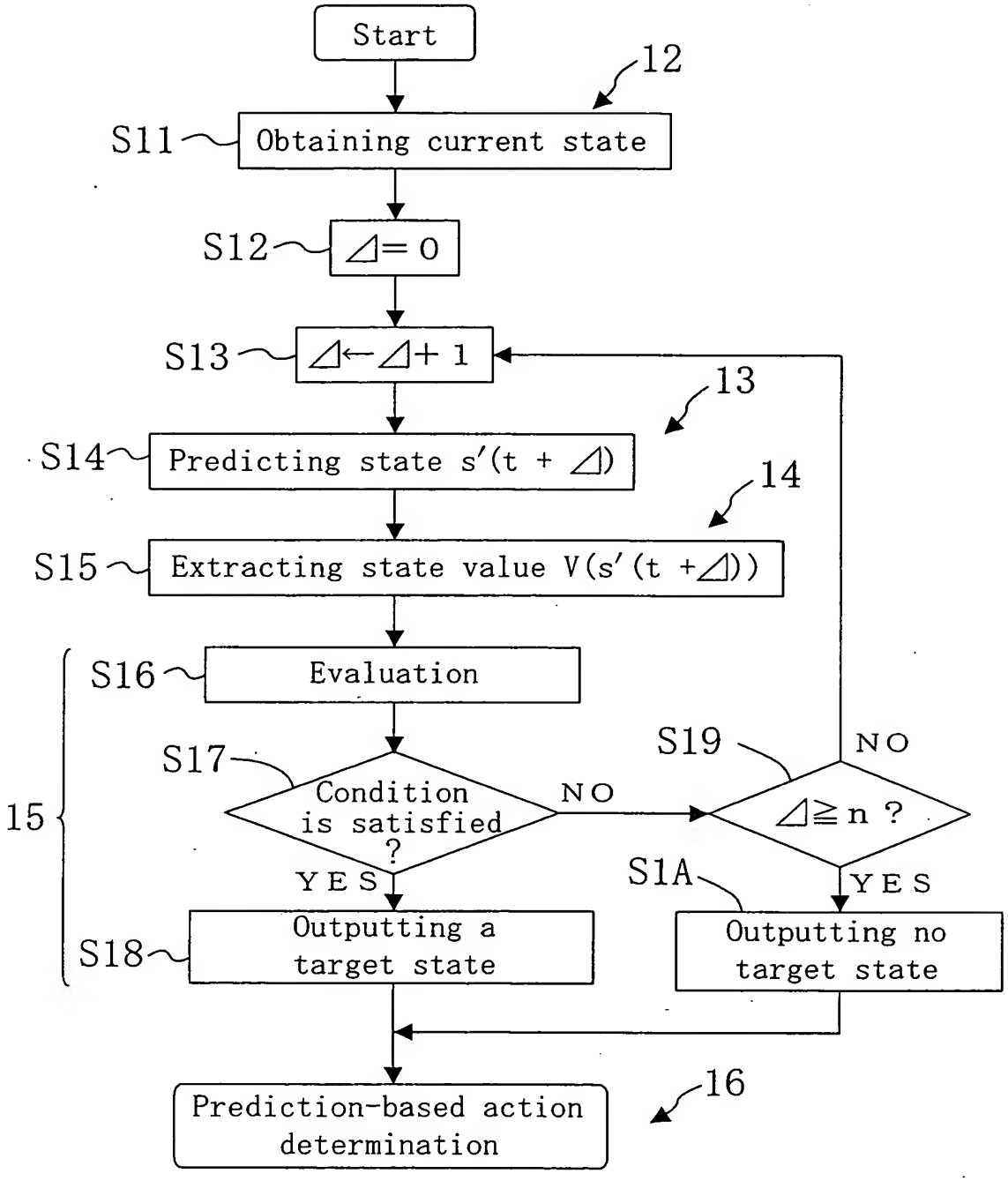

Start
12
S11
Obtaining current state
S12
⊿= 0
S13
⊿←⊿+ 1
13
S14
Predicting state s'(t + ⊿)
14
S15
Extracting state value V(s'(t +⊿))
S16
Evaluation
S17
Condition is satisfied ?
NO
S19
⊿≧n ?
NO
15
YES
YES
S1A
S18
Outputting a target state
Outputting no target state
16
Prediction-based action determination

FIG. 6

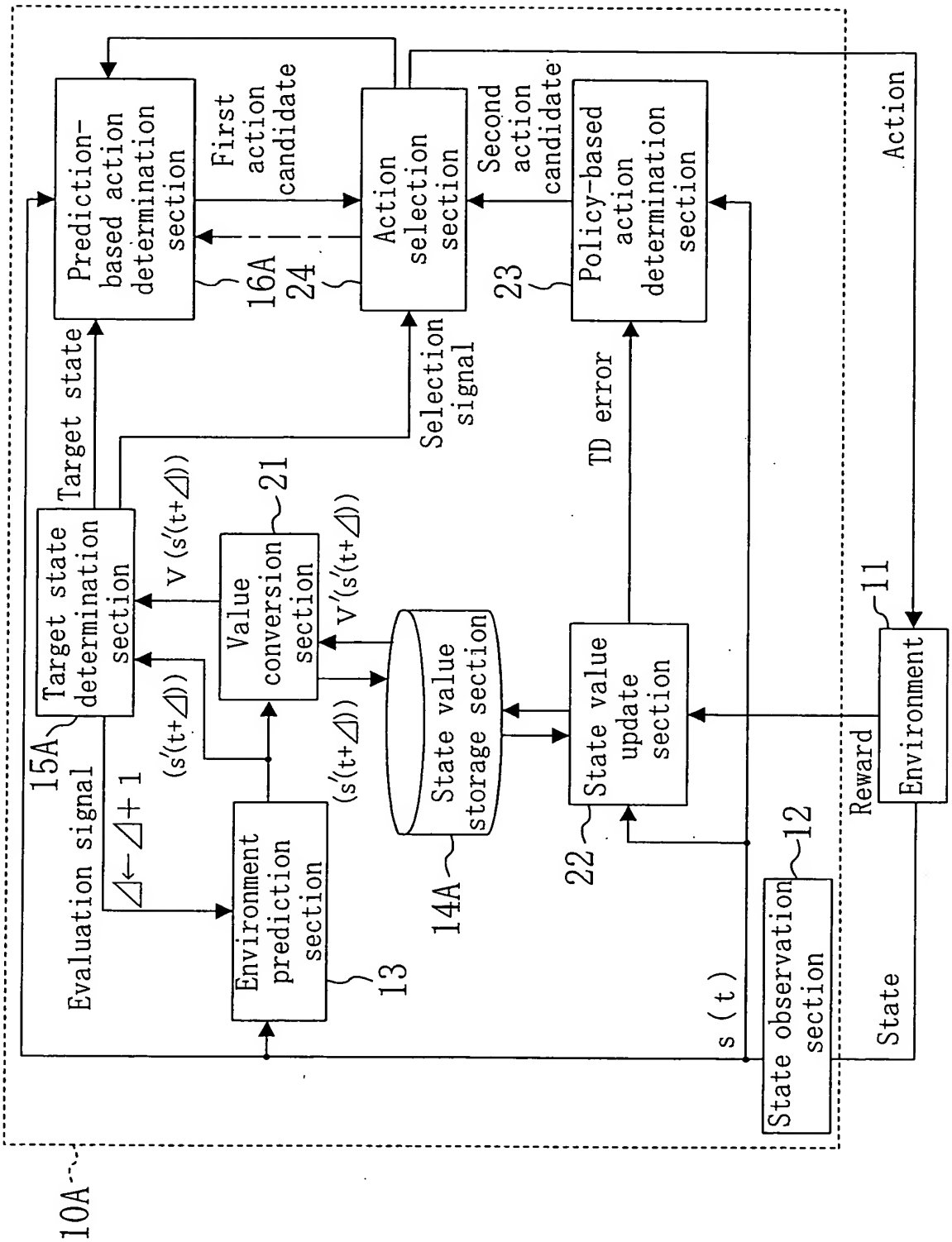

10A
Prediction-based action determination section
16A
First action candidate
Action selection section
24
Second action candidate
Policy-based action determination section
23
Action
Target state
Selection signal
TD error
Target state determination section
15A
Evaluation signal
Δ ← Δ + 1
V (s'(t+Δ))
Value conversion section
21
V'(s'(t+Δ))
(s'(t+Δ))
(s'(t+Δ))
State value storage section
14A
State value update section
22
11
Environment
Reward
Environment prediction section
13
s (t)
State observation section
12
State

FIG. 7

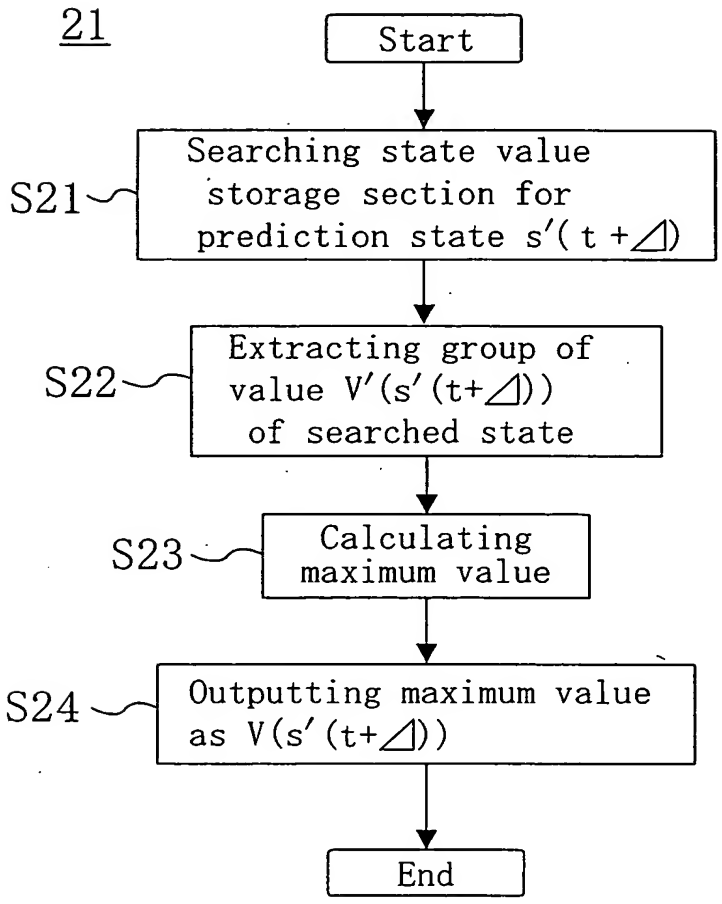

21
Start
S21
Searching state value storage section for prediction state s′(t +⊿)
S22
Extracting group of value V′(s′(t+⊿)) of searched state
S23
Calculating maximum value
S24
Outputting maximum value as V(s′(t+⊿))
End

FIG. 8

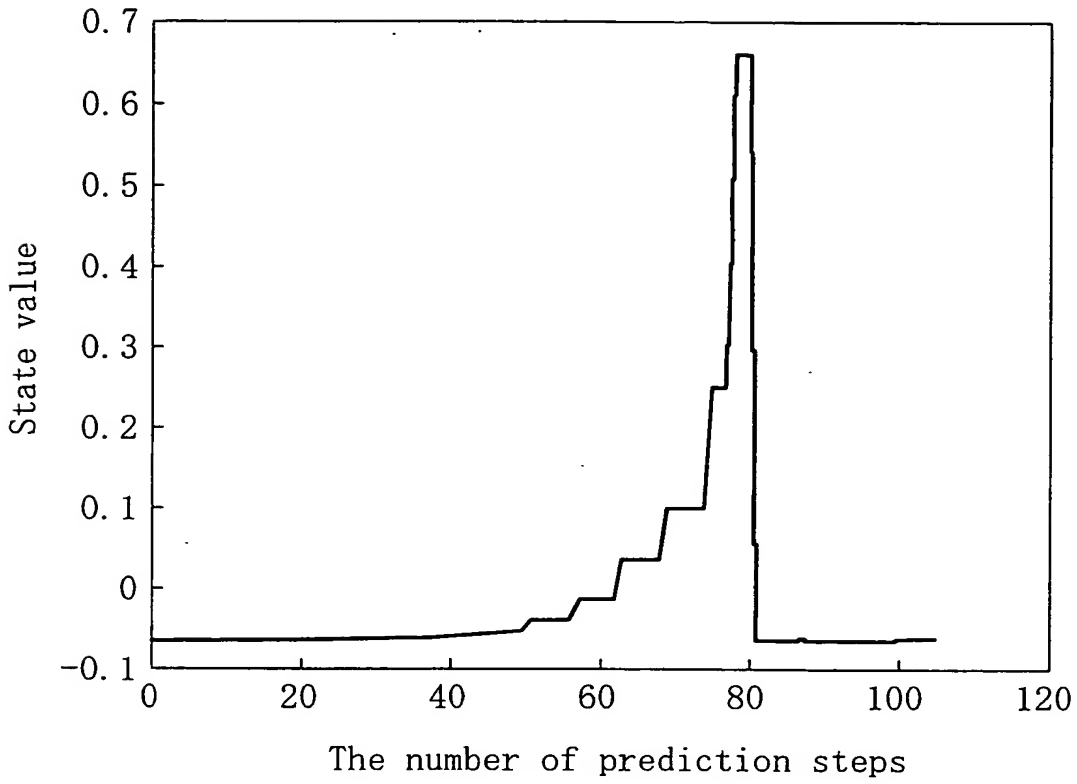

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
State value
0
20
40
60
80
100
120
The number of prediction steps

FIG. 9

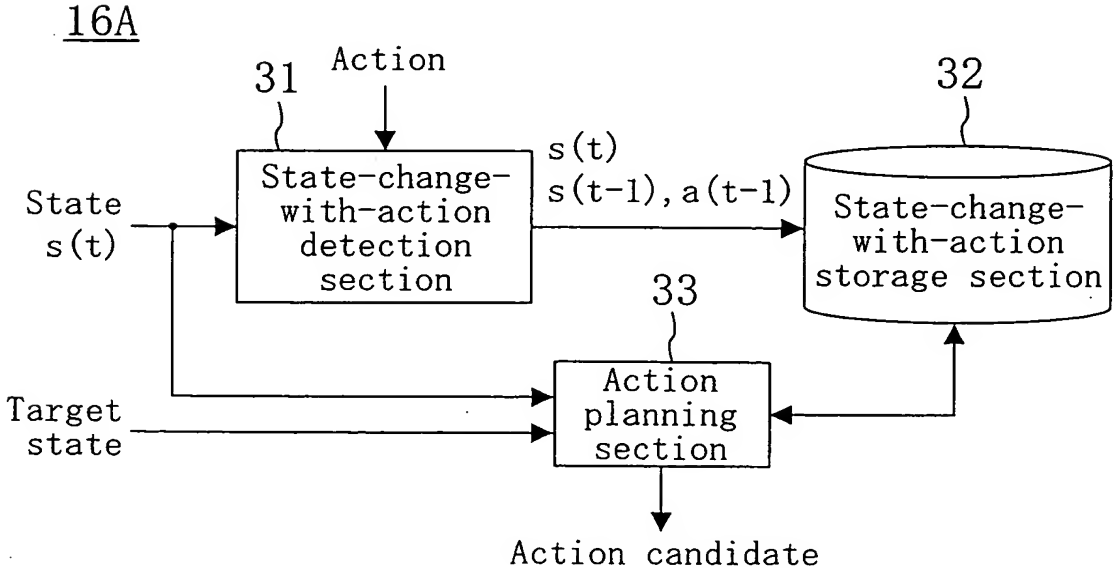

16A
31
Action
32
State
s(t)
State-change-with-action detection section
s(t)
s(t-1), a(t-1)
State-change-with-action storage section
33
Target state
Action planning section
Action candidate

FIG. 10

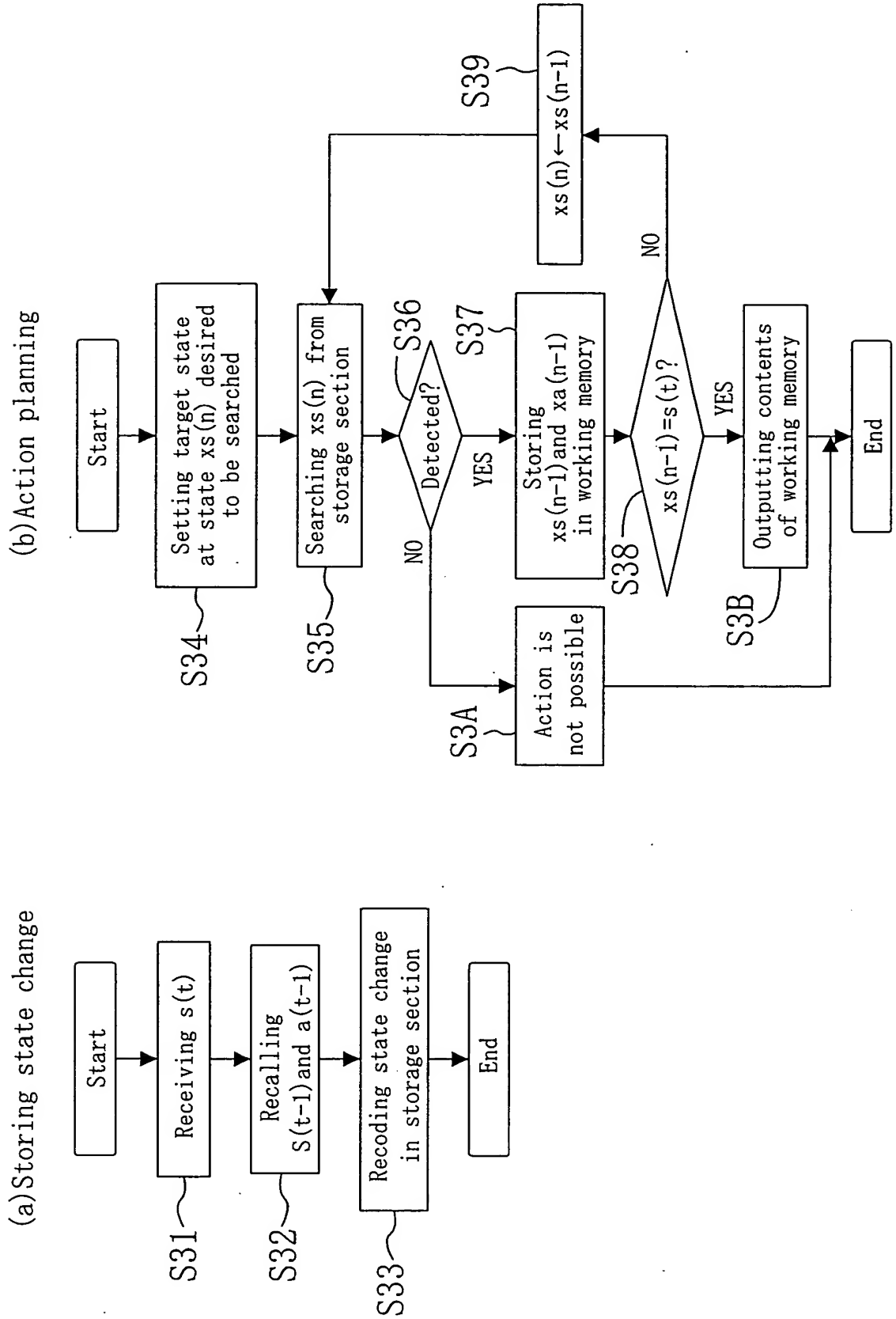

(a) Storing state change
Start
Receiving s(t)
Recalling s(t-1) and a(t-1)
Recoding state change in storage section
End
S31
S32
S33
(b) Action planning
Start
Setting target state at state xs(n) desired to be searched
Searching xs(n) from storage section
Detected?
YES
NO
Storing xs(n-1) and xa(n-1) in working memory
xs(n-1)=s(t)?
YES
NO
xs(n)←xs(n-1)
Outputting contents of working memory
End
Action is not possible
S34
S35
S36
S37
S38
S39
S3A
S3B

FIG. 11

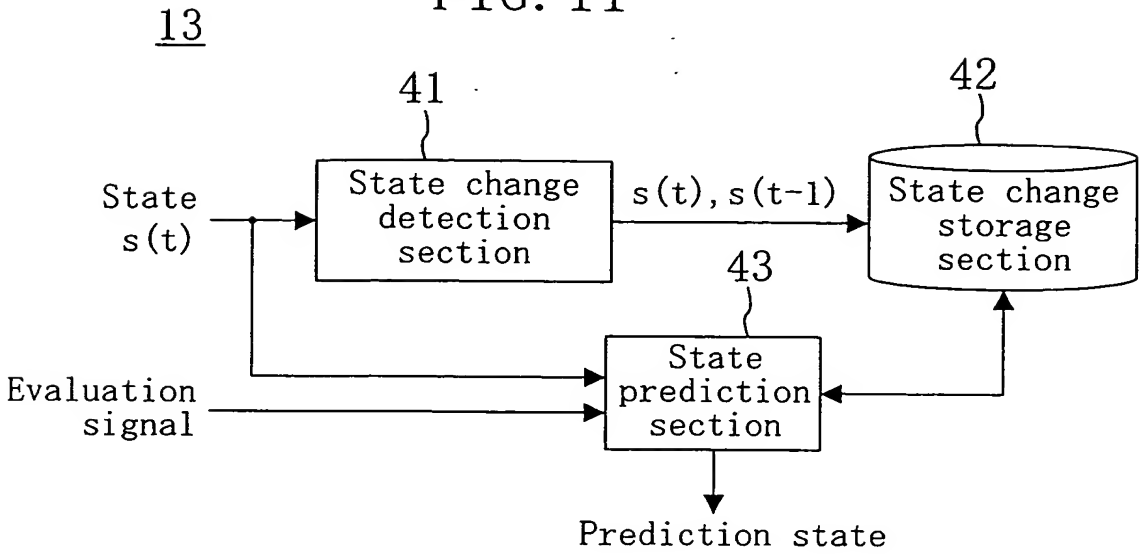

13
41
42
State
s(t)
State change detection section
s(t), s(t-1)
State change storage section
43
Evaluation signal
State prediction section
Prediction state

FIG. 12

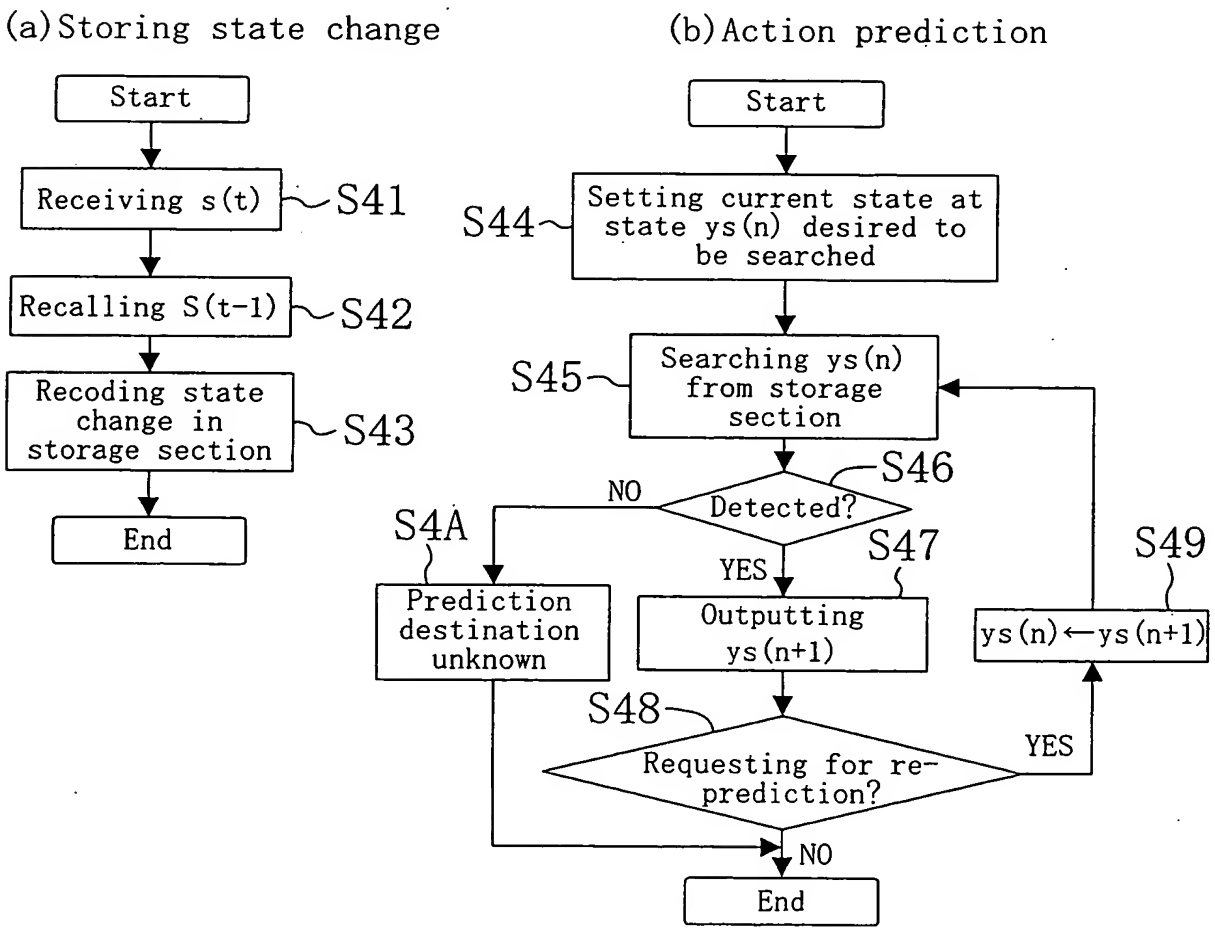

(a) Storing state change
Start
Receiving s(t)
S41
Recalling S(t-1)
S42
Recoding state change in storage section
S43
End
(b) Action prediction
Start
S44
Setting current state at state ys(n) desired to be searched
S45
Searching ys(n) from storage section
S46
Detected?
NO
YES
S4A
Prediction destination unknown
S47
Outputting ys(n+1)
S49
ys(n)←ys(n+1)
S48
Requesting for re-prediction?
YES
NO
End

FIG. 13

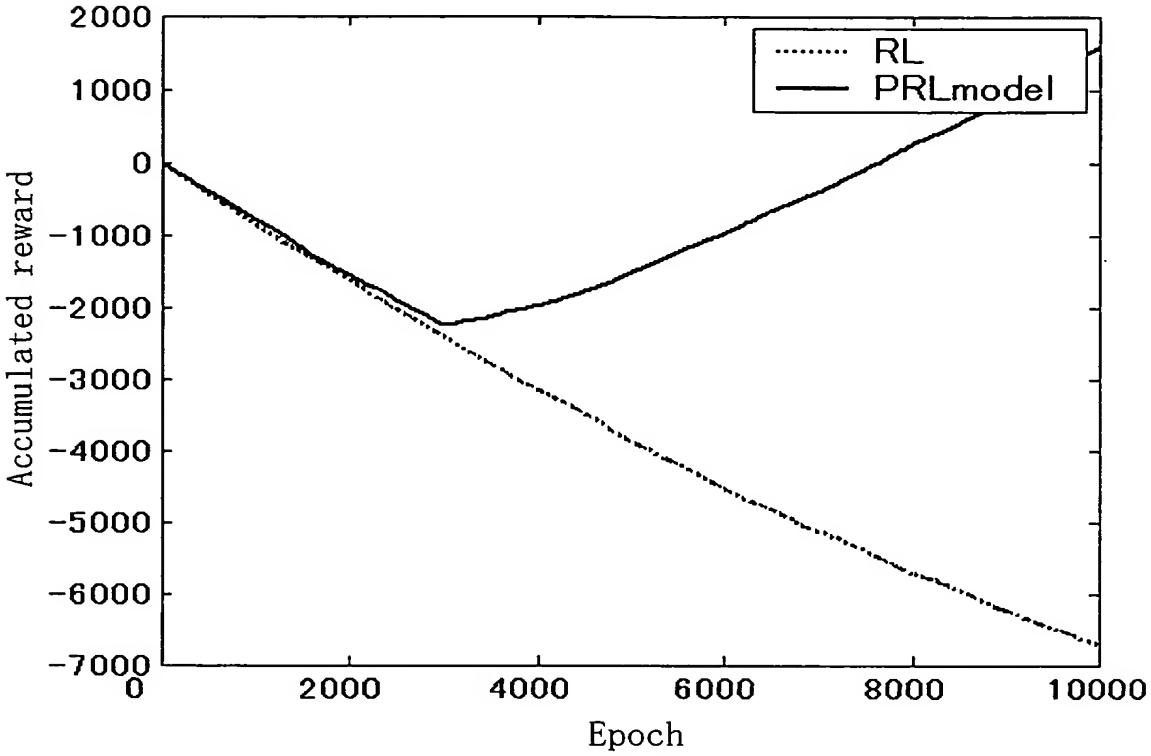

RL
PRLmodel
Accumulated reward
2000
1000
0
-1000
-2000
-3000
-4000
-5000
-6000
-7000
0
2000
4000
6000
8000
10000
Epoch

FIG. 14

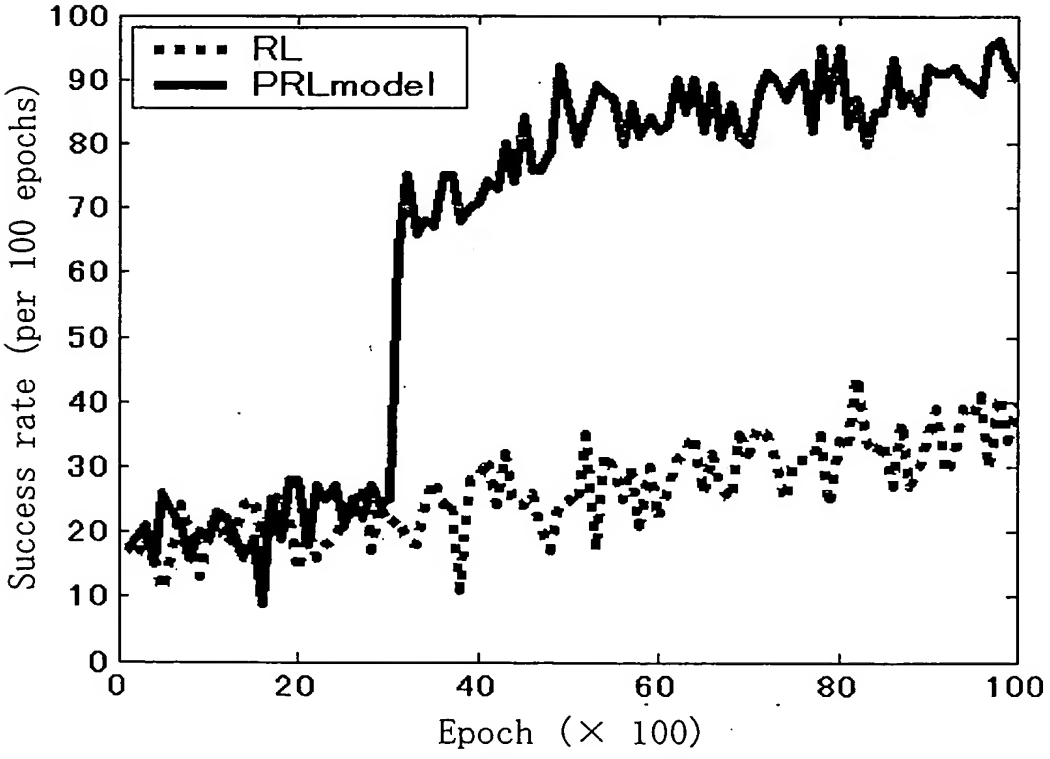

RL
PRLmodel
Success rate (per 100 epochs)
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
Epoch (× 100)

FIG. 15

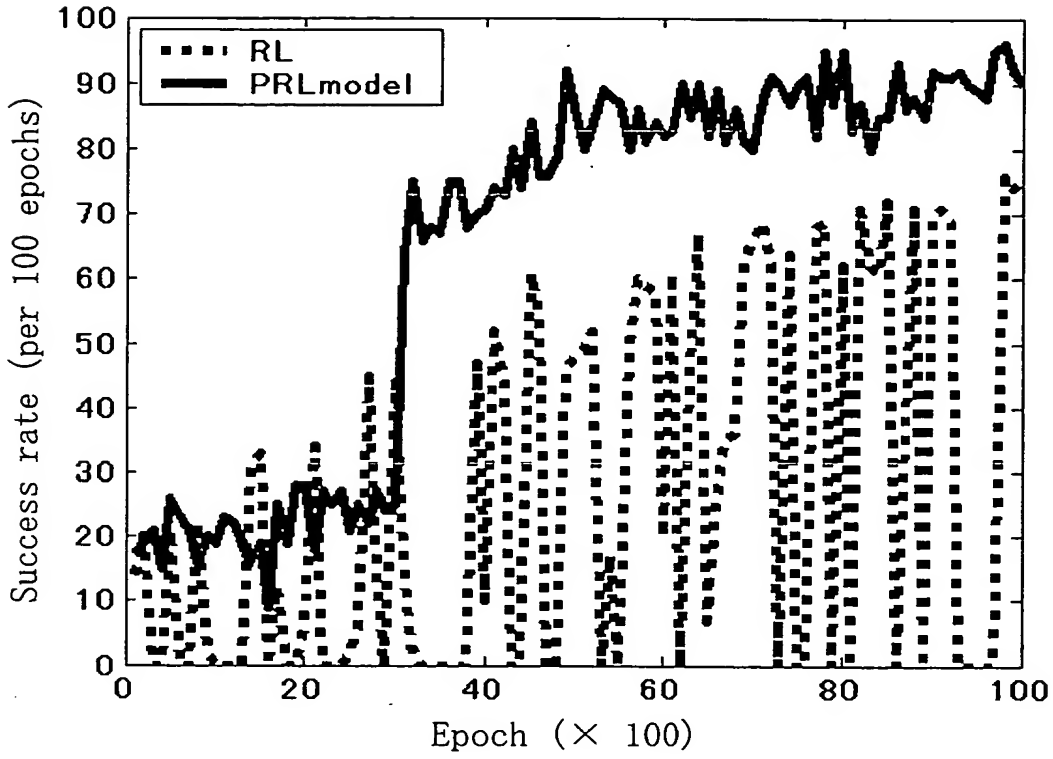

RL
PRLmodel
Success rate (per 100 epochs)
Epoch (× 100)
0
10
20
30
40
50
60
70
80
90
100
0
20
40
60
80
100

FIG. 16

(a) Before learning

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(b) Middle of learning

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3

(c) After completion of learning

1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8

FIG. 17

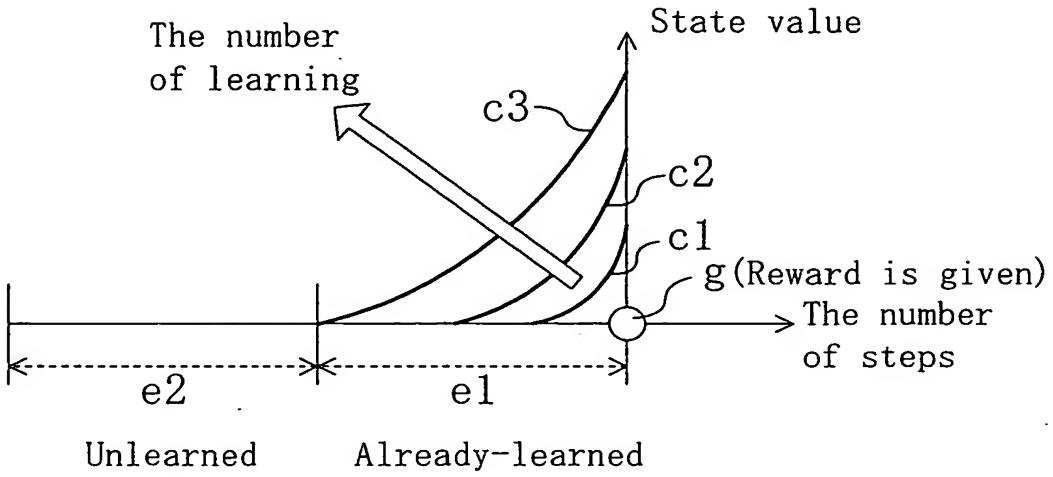

The number
of learning
State value
c3
c2
c1
g(Reward is given)
The number
of steps
e2
e1
Unlearned
Already-learned

FIG. 18

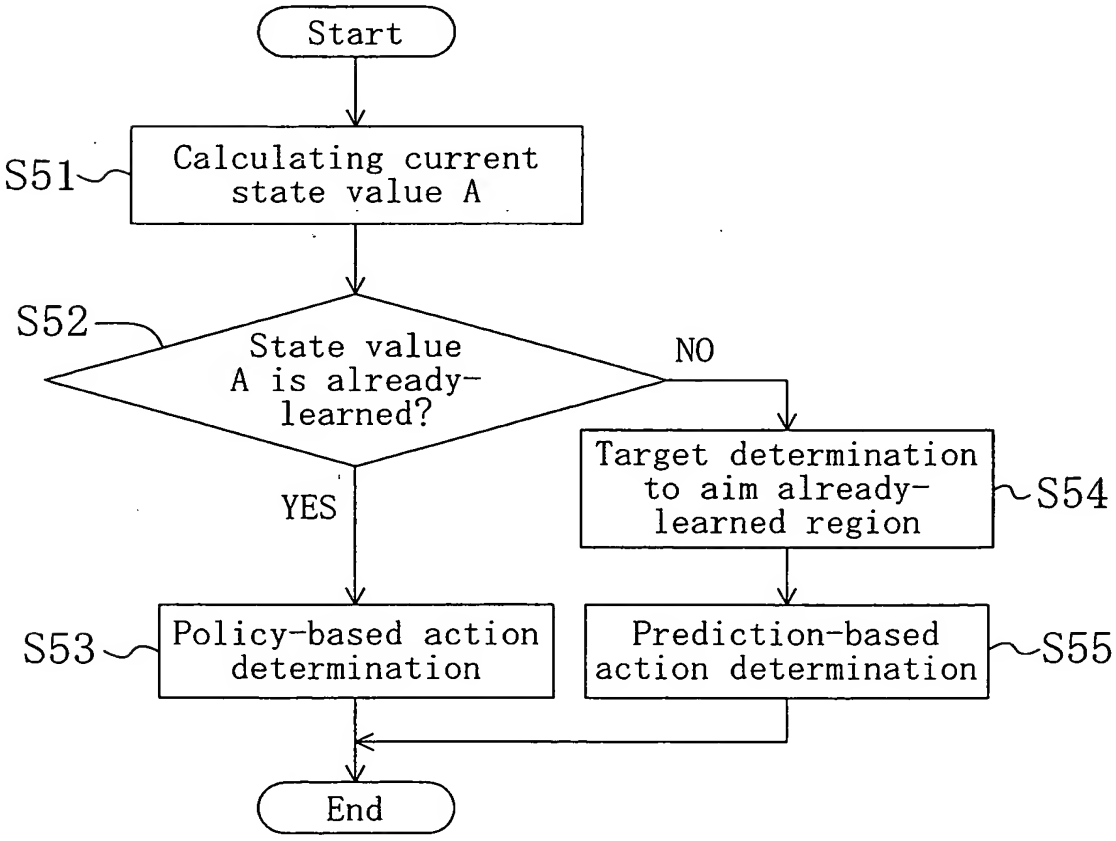
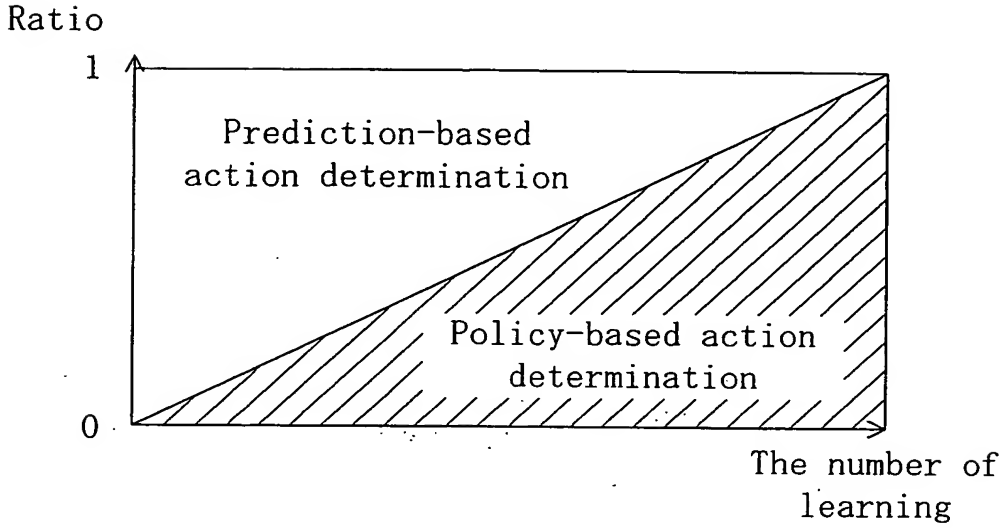

Start
S51
Calculating current state value A
S52
State value A is already-learned?
NO
YES
Target determination to aim already-learned region
S54
S53
Policy-based action determination
Prediction-based action determination
S55
End

FIG. 19


Ratio
1
0
Prediction-based
action determination
Policy-based action
determination
The number of
learning